PHONETIC REDUCTION, VOWEL DURATION, AND PROSODIC STRUCTURE

Rachel Steindel Burdin, Cynthia G. Clopper

The Ohio State University
burdin.1@osu.edu, clopper.1@osu.edu

ABSTRACT

Word frequency, phonological neighborhood density, semantic predictability in context, and discourse mention have all been previously found to cause reduction of vowels. Other researchers have suggested that reduction based on these factors is reflective of a unified process in which "redundant" or "predictable" elements are reduced, and that this reduction is largely mediated by prosody. Using a large read corpus, we show that these four factors show different types of reduction effects, and that there are reduction effects of prosody independent of duration, and vice versa, suggesting the existence of multiple processes underlying reduction.

Keywords: phonetic reduction, prosodic reduction, vowel reduction, predictability

1. INTRODUCTION

The production of vowels exhibits considerable intra-talker variation. This variation has been found to be conditioned by many different factors, including lexical properties such as word frequency [10] and phonological neighborhood density [10], and properties of the discourse such as semantic predictability in the sentence context [8, 9] and mention in the discourse context [7, 2]. Some degree of commonality between these and other factors has been proposed: Aylett and Turk [1] classified word frequency, trigram word probability, and discourse mention as "redundancy factors"; Baker and Bradlow [2] described both high frequency words and second mentions as "probable"; and Turnbull [14] described non-first mention items, focused items, and discourse-predictable items as "predictable".

To explain why word frequency, trigram word probability, and discourse mention all condition vowel reduction, Aylett and Turk [1] proposed the Smooth Signal Redundancy Hypothesis (SSRH), which states that these "redundancy factors" affect vowel reduction via prosody. The underlying assumption of the SSRH is that information transfer in communication should be relatively consistent. Less contextually redundant elements should be phonetically or prosodically enhanced, whereas more redundant elements should be reduced.

Aylett and Turk examined effects of redundancy on syllable duration, and argued that reduction in duration based on these redundancy factors is largely mediated by prosody. However, their results suggest that other factors besides prosody may play a role in facilitating reduction, as the prosodic factors they controlled for only accounted for approximately 60% of the variance observed in duration. In addition, Baker and Bradlow [2] found second mention reduction even when controlling for the presence or absence of a pitch accent, suggesting that there are multiple factors, including prosody, which affect vowel reduction. Finally, Calhoun's [6] model of the prosodic signaling of information structure includes multiple additional factors beyond pitch accenting, including metrical constraints, and overall accentability of a word. Together these results suggest that the strong version of the SSRH, in which vowel reduction is mediated entirely through the prosodic structure, does not hold.

Further, it is unclear whether the above factors (i.e., word frequency, neighborhood density, semantic predictability in context, and discourse mention) which are known to condition vowel reduction, are truly reflective of the same underlying process or processes. Munson and Solomon [10] found that neighborhood density and word frequency exhibit independent effects on vowel production, suggesting additive effects of different sources of redundancy or predictability. Burdin et al. [5] found that word frequency and neighborhood density interact differently with discourse-level factors such as speaking style and discourse mention, suggesting a difference in the underlying processes leading to phonetic reduction.

The current study explores two questions raised by the research described above, using a large corpus of read speech. The first question is to what extent vowel reduction based on the above factors is mediated exclusively by the prosodic structure, including pitch accenting and phrasing. The second question is how word frequency, neighborhood den-

sity, predictability in context, and discourse mention affect vowel reduction and are thus reflective, or not, of the same underlying process or processes.

2. METHODOLOGY

2.1. Stimulus materials

A set of 30 short stories was created, modeled after those used by [2]. A sample story is given in (1).

(1) During her summer internship, Jessica gave up her hopes of becoming a veterinary assistant after seeing *black* pus *ooze* out of the *sheep's* infected *leg*. Although she'd seen the size of the growth on the **sheep's leg**, she was altogether unprepared to see it **ooze** and *bleed* an inky **black** fluid. As she watched the animal **bleed** while the veterinarian tended to it, Jessica realized she definitely wasn't cut out for this!

The stories contained a total of 235 target words. Each target word appeared in a story twice: in the story in (1), the first mention of each word is in italics; the second mention in bold. The frequency and phonological neighborhood density of each target word was calculated using the Hoosier Mental Lexicon [11]. Care was taken to ensure that frequency and density varied independently. Although some research on vowel reduction uses a binary classification between "easy" words, which are high frequency and have a low neighborhood density, and "hard" words, which are low frequency and have a high neighborhood density (e.g., [15]), in this study, all four combinations of low and high frequency and low and high density were included. For example, target words like *bleed*, which are low frequency and have a low neighborhood density, and target words like *leg*, which are high frequency and have a high neighborhood density, were also included. The predictability of each word in its sentence context was determined using a separate cloze task. Each target word appeared in both more and less predictable contexts. For example, based on the results of the cloze task, the first mention of *ooze* (in "black pus ooze") is more predictable in context than the second mention of *ooze* (in "ooze and bleed"). Like the lexical factors, the predicability and discourse mention factors were manipulated independently, so that across target words, both first and second mention appeared in both low and high predictability contexts.

Participants were instructed to read each story as if they were talking to a friend to elicit plain lab speech. These instructions for speaking style were presented before each story as a reminder. The stories were presented in a different random order for each participant.

Data from 10 participants were analyzed in the current study. The participants were all undergraduates at a large university in the midwestern United States and received either partial course credit or $10 for participating. The sound files for each story were forced-aligned to the original story text using the Penn Forced Aligner [16]. Vowel boundaries for the target words were then hand-corrected.

2.2. Prosodic annotation

To explore the effects of prosody on phonetic vowel reduction, prosodic annotation of the read stories was conducted using the ToBI (Tones and Break Indices) annotation system [3] by the first author and two trained undergraduate research assistants. For each target word, the presence or absence of a pitch accent, as well as pitch accent type was coded, along with the strength of the preceding and following prosodic breaks, and where applicable, the type of phrase accent and/or phrase accent and boundary tone associated with a following prosodic break of strength 3 or 4. Annotations were reviewed in group meetings; annotations for which a coder was unsure or for which there was disagreement among the coders were discussed until an annotation could be agreed upon. A small number of tokens on which consensus could not be reached were excluded from the analysis, along with disfluencies. Words with an adjacent level 2 or 0 break were also excluded, as a 2 break indicates a certain degree of mismatch between the phonetic cues and the perceived prosodic juncture, and a 0 break indicates that the target word had some degree of cliticization with the preceding or following word, both of which indicate a certain degree of disfluent or unusual prosodic structure around the word. In addition, if a word was produced with a disfluency in its first mention, both mentions were excluded from the analysis. Of the 4,700 target word tokens (10 talkers x 235 target words x 2 mentions), 442 tokens were excluded at this stage.

The annotations were completed using Praat [4]. The duration of the target vowels and prosodic annotations were automatically extracted from the TextGrids. Tokens whose vowel duration was more than three standard deviations from the talker's mean vowel duration were excluded from the analysis (462 tokens), leaving 3,796 tokens for analysis. Due to the low number of L* and L*+H pitch accents, pitch accent distinctions were simplified into a

two category distinction of rising (L*+H and L+H*) and non-rising (H*, !H*, and L*) pitch accents.

3. RESULTS

To assess the effects on duration of the linguistic factors contributing to phonetic reduction independently of pitch accenting and phrasing, linear mixed effects models were built predicting vowel duration, with log frequency, neighborhood density, cloze predictability, and mention, as well as pitch accent type (unaccented, rising, or non-rising) and following break strength (word level, intermediate phrase, or intonational phrase) as fixed effects. Random intercepts for talker and target word were also included in the model to control for variability in the length of words and talker speaking rates. Models were also built with various levels of interactions between the fixed effects; however, these models failed to converge, so the model without interactions between the fixed effects is reported and interpreted. Significance was assessed at $|t| > 2.0$.

Vowels in high density words were longer than vowels in low density words ($\beta = 0.0009$, $t = 4.917$). In addition, as expected given previous research on prosodic structure and duration, vowels in words with a non-rising ($\beta = 0.0087$, $t = 6.928$) or rising pitch accent ($\beta = 0.0101$, $t = 4.637$) were longer than vowels in unaccented words, and vowels in words that were followed by either a intermediate phrase ($\beta = 0.022$, $t = 14.208$) or intonational phrase ($\beta = 0.026$, $t = 15.924$) break were longer than vowels in words followed by a word level break. The independent effect of density on duration was small relative to the effects of pitch accenting and phrasing, as can be seen by the size of the model coefficients, but significant. This result suggests that phonetic vowel reduction cannot be attributed entirely to effects of prosody.

To further explore the relationships among prosodic structure, phonetic reduction factors, and duration, models were also built to predict pitch accenting from the linguistic factors contributing to phonetic reduction. A logistic mixed effects model was built predicting pitch accenting (pitch accented or not), with log frequency, neighborhood density, cloze predictability, mention, and presence or absence of a following intermediate or intonational phrase boundary as fixed effects, duration as a covariate, and random intercepts for talkers and target words. Models with various levels of interactions between the fixed effects failed to converge, so the model without interactions is reported and interpreted.

Less frequent ($\beta = -0.66$, $z = -5.277$, $p < 0.001$) and less predictable ($\beta = -1.39$, $z = 3.920$, $p < 0.001$) words were more likely to be produced with a pitch accent than more frequent and more predictable words, consistent with Aylett and Turk's [1] suggestion that "redundant" elements are prosodically reduced. Somewhat trivially, pitch-accented words were also longer ($\beta = 18.03$, $z = 8.697$, $p < 0.001$) than non-pitch-accented words, as demonstrated in the previous analysis, and were more likely to appear before a word-level break (level 1 break) than a phrase-level break (level 3 or 4) ($\beta = 0.664$, $z = 3.713$, $p < 0.001$).

Tables 1 and 2 illustrate the effects of predictability and frequency on pitch accenting, respectively. To clarify the presentation of the results a boundary was defined between high predictability (cloze predictability > 0.125) and low predictability (cloze predictability < 0.125) words, as well as between high frequency (log frequency > 2.5) and low frequency (log frequency < 2.5) words.

An inspection of Table 1 reveals that 77% of the high predictability words were accented, compared to 86% of the low predicability words; however, the breakdown of pitch accent type between high and low predictability words is nearly identical (93% non-rising/ 7% rising in both cases). Likewise, as seen in Table 2, there is a large difference between the number of high frequency (75%) and low frequency (86%) words which were pitch accented, but the breakdown of pitch accent types between low and high frequency words is very similar.

Table 1: Predictability and pitch accenting

	High predictability		Low predictability	
Unaccented	441	(23%)	225	(14%)
Accented	1471	(77%)	1554	(86%)
Non-rising	1368	(93%)	1444	(93%)
Rising	104	(7%)	110	(7%)

Table 2: Frequency and pitch accenting

	High frequency		Low frequency	
Unaccented	432	(25%)	259	(14%)
Accented	1298	(75%)	1724	(86%)
Non-rising	1216	(94%)	1592	(92%)
Rising	82	(6%)	132	(8%)

To explore these effects of phonetic reduction on

pitch accent type, multinominal logistic mixed effects models were built to predict pitch accent type (H*, !H*, L*, L+H*, L*+H) from the linguistic factors contributing to phonetic reduction; however, these models failed to converge. Thus, as in the duration analysis, pitch accent type was reduced to a two-way contrast between non-rising (H*, !H*, and L*) and rising (L+H* and L*+H) pitch accents and a logistic mixed effects model was built. Log frequency, neighborhood density, cloze predictability, mention, and presence or absence of a following intermediate or intonational phrase boundary were included as fixed effects, with duration as a covariate, and with random intercepts for talkers and target words. As in the previous analyses, models with various levels of interactions between the fixed effects failed to converge, so the model without interactions is reported and interpreted.

First mention words were more likely to be produced with a rising pitch accent (β= -0.497, z = -3.156, $p < 0.01$) than second mention words. This result is consistent with the use of rising pitch accents, especially L+H*, for focused items [12]. Non-rising pitch accents were more likely to appear before a word level break than an intermediate or intonational phrase break (β=0.46, z = 2.708, $p < 0.01$); that is, more of target words produced with a H*, !H*, or L* were produced without a following phrase break.

Table 3 illustrates the effect of discourse mention on pitch accenting. For this variable, the overall presence of pitch accenting was similar for first and second mentions (83% and 82%); however, more first mention words were produced with rising pitch accents (8%) than second mention words (6 %).

Table 3: Mention and pitch accenting

	First mention		Second mention	
Unaccented	329	(17%)	362	(18%)
Accented	1526	(83%)	1496	(82%)
Non-rising	1398	(92%)	1410	(94%)
Rising	128	(8%)	86	(6%)

4. DISCUSSION

Frequency, neighborhood density, semantic predictability, and discourse mention all had reduction effects. These findings were largely expected, given previous research; however, this study found a duration effect for density, which has not been found in previous studies (e.g., [10]), or, has been found and attributed to segmental effects ([13]).

Each of these different factors had different types of reduction effects. When prosodic factors (pitch accenting and presence/absence of a following phrase break) were controlled for in the analysis of vowel duration, words with low neighborhood density still exhibited a phonetic reduction effect, although none of the other linguistic factors were significant. Similarly, when vowel duration was controlled for, high predictability words were less likely to be pitch accented than low predictability words and high frequency words were less likely to be pitch accented than low frequency words. Finally, when vowel duration was controlled for, second mention words exhibited differences in pitch accent type compared to first mention words, with first mention words being more likely to have rising pitch accents than second mention words.

These independent effects of duration and prosodic structure suggest that temporal reduction (i.e., vowel duration) based on frequency, neighborhood density, semantic predictability, and discourse mention is not entirely mediated by prosodic factors, contra the strong version of Aylett and Turk's [1] theory, and consistent with previous work that found that prosody was largely, but not entirely, responsible for temporal reduction [1, 2, 6]. Prosodic structure exhibits independent effects on duration and phonetic reduction exhibits independent effects on prosodic structure. In addition, the fact that the four phonetic reduction factors exhibited different patterns of reduction – predictability and frequency through pitch-accenting, mention through pitch accent type, and density through duration – suggests that there are multiple processes underlying phonetic reduction for these factors, rather than one unified process related to predictability or redundancy.

5. ACKNOWLEDGEMENTS

The authors would like to thank Abby Walker for help with stimulus design; McKenna Reeher and Shannon Melvin for help in data collection; Anna Crabb, Rachel Monnin, Sarah Mabie, Christine Pretchel, Shannon Melvin, and Erin Luthern for help with data annotation; and Rory Turnbull for help with data collection and annotation, and his comments. This work was supported by NSF Grant BCS-1056409.

6. REFERENCES

[1] Aylett, M., Turk, A. 2004. The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic promi-

nence and duration in spontaneous speech. *Language and Speech* 47(1), 31 – 56.

[2] Baker, R. E., Bradlow, A. R. 2009. Variability in word duration as a function of probability, speech style, and prosody. *Language and Speech* 52(4), 391 – 413.

[3] Beckman, M. E., Ayers Elam, G. 1997. Guidelines for ToBI labelling, version 3. the Ohio State University Research Foundation. www.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdfwww.ling.ohio-state.edu/~tobi/ame_tobi/labelling_guide_v3.pdf.

[4] Boersma, P., Weenink, R. 2014. Praat: doing phonetics by computer. http://www.praat.org.

[5] Burdin, R. S., Turnbull, R., Clopper, C. G. 2014. Interactions among lexical and discourse characteristics in vowel production. *Journal of the Acoustic Society of America* 136, 2172.

[6] Calhoun, S. 2010. The centrality of metrical structure in signaling information structure: a probabilistic perspective. *Language* 82(1), 1 – 42.

[7] Fowler, C. A., Housum, J. 1987. Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26, 489 – 504.

[8] Kalikow, D. N., Stevens, K. N., Elliott, L. L. 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustic Society of America* 61, 1337 – 1351.

[9] Lieberman, P. 1963. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech* 6, 172 – 187.

[10] Munson, B., Solomon, N. 2004. The effect of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research* 47, 1048 – 1058.

[11] Nusbaum, H. C., Pisoni, D. B., Davis, C. L. 1984. Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. In: *Research on speech perception progress report no. 10*. Bloomington, IN: Speech Research Laboratory, Indiana University 357 – 376.

[12] Pierrehumbert, J., Hirschberg, J. 1990. The meaning of intonational contours in the interpretation of discourse. In: Cohen, P., Morgan, J., Pollack, M., (eds), *Intentions in Communication*. MIT Press 271 – 311.

[13] Scarborough, R. 2012. Lexical similarity and speech production: neighborhoods for nonwords. *Lingua* 112, 164 – 176.

[14] Turnbull, R. The role of predictabiltiy in intonational variability. Under review.

[15] Wright, R. A. 2004. Factors of lexical competition in vowel articulation. In: Local, J., Ogden, R., Temple, R., (eds), *Phonetic interpretation: Papers in Laboratory Phonology*. Cambridge, U. K.: Cambridge University Press 75 – 87.

[16] Yuan, J., Liberman, M. 2008. Speaker identification on the SCOTUS corpus. *Journal of the Acoustic Society of America* 123(5), 3878.